nature structural & molecular biology

Structural basis for DNA bridging by barrier-to-autointegration factor

Christina Marchetti Bradley, Donald R Ronning, Rodolfo Ghirlando, Robert Craigie & Fred Dyda

The ability of barrier-to-autointegration factor (BAF) to bind and bridge DNA in a sequence-independent manner is crucial for its role in retroviral integration and a variety of cellular processes. To better understand this behavior, we solved the crystal structure of BAF bound to DNA. The structure reveals that BAF bridges DNA using two pairs of helix-hairpin-helix motifs located on opposite surfaces of the BAF dimer without changing its conformation.

Several viral and cellular proteins are important in retroviral integration, ensuring the stable insertion of the retroviral cDNA into the host genome[1]. The cellular protein BAF is a stable component of human immunodeficiency virus (HIV)[2] and Moloney murine leukemia virus (MoMLV)[3] preintegration complexes (PICs) that stimulates PIC activity[4] and specifically protects the MoMLV PIC from autointegration (the suicidal insertion of the cDNA back into itself)[5]. BAF is highly conserved in metazoans and has important roles in chromosomal organization, nuclear assembly and gene expression[6]. Knockdown of BAF by RNA-mediated interference in *Caenorhabditis elegans*[7,8] and a BAF null mutant in *Drosophila melanogaster*[9] are lethal at the larval stage, showing that BAF is an essential protein. It is a compact, all-helical obligatory dimer[10,11]. Each monomer contains two copies of a nonspecific DNA-binding motif, the helix-hairpin-helix (HhH)[12,13] motif[10], that is found in over 100 proteins that bind DNA nonspecifically[12,13].

BAF has a unique ability to bind and bridge double-stranded (ds) DNA in a sequence-independent manner[5,7]. For example, with 21 base pairs (bp) of DNA, BAF forms a discrete higher-order complex consisting of six DNA molecules and six BAF dimers[7]. Although these data firmly establish the DNA-bridging property of BAF, they do not show the structural organization of the bridged complex or explain how BAF interacts with longer DNA. BAF may bridge distant regions within a DNA molecule, resulting in a collapsed chain. Alternatively, BAF may oligomerize upon binding DNA, consistent with reports that BAF undergoes a DNA-induced conformational change[6]. To determine how BAF achieves high-affinity, nonspecific DNA binding and bridging, we characterized complexes of human BAF and short DNA and solved the crystal structure of a BAF–DNA complex (**Supplementary Methods** online).

With 7 bp of DNA, BAF formed a discrete complex consisting of a single BAF dimer and two DNA molecules (**Supplementary Figs. 1** and **2**, **Supplementary Table 1** and **Supplementary Discussion** online). This shows that BAF does not necessarily oligomerize to a hexameric species upon binding DNA and suggests that the structure of the BAF–7-bp DNA complex may represent the fundamental BAF-DNA interaction. In contrast, with 9 and 11 bp of DNA, BAF formed a heterogeneous mixture of 1:2 and higher-order complexes of indiscriminate stoichiometry (**Supplementary Figs. 2** and **3**, **Supplementary Table 1** and **Supplementary Discussion** online), suggesting

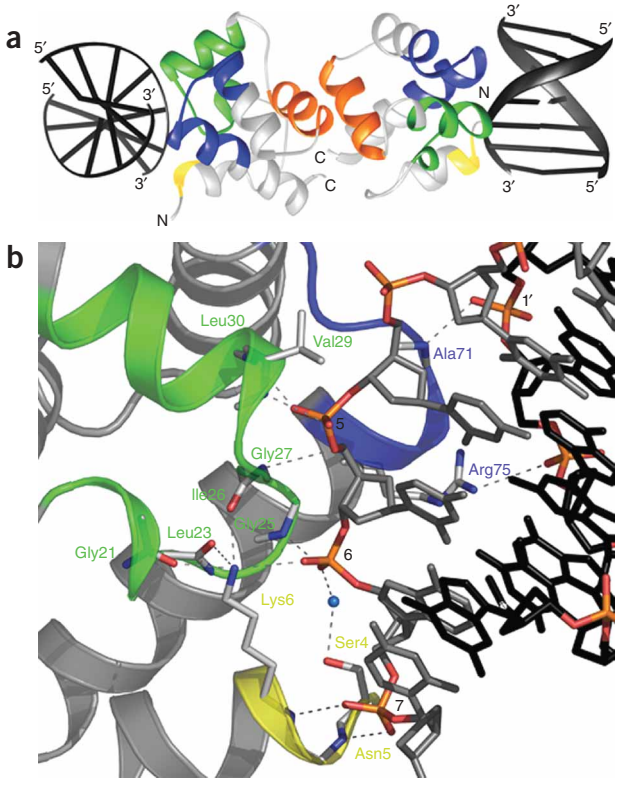

Figure 1 Crystal structure of BAF–7-mer dsDNA complex. (**a**) Overall structure. The HhH (green) and pseudo HhH (blue) motifs along with the connecting helix (orange) comprise a five-helix $(HhH)_2$ motif[13]. Yellow, interacting region of α1; black, DNA; N and C, protein termini. (**b**) Detailed view of BAF-DNA interactions. Ribbons and residue labels are colored as in **a**, with sticks showing important residues (blue, nitrogen; red, oxygen; gray, carbon). DNA strands are gray and black with the phosphates in orange, and nucleotides are numbered. Dotted lines indicate hydrogen bonds.

Laboratory of Molecular Biology, National Institute of Diabetes & Digestive & Kidney Diseases, US National Institutes of Health (NIH), 5 Center Drive, Bethesda, Maryland 20892, USA. Correspondence should be addressed to F.D. (fred.dyda@nih.gov).

Received 11 May; accepted 15 August; published online 11 September 2005; doi:10.1038/nsmb989

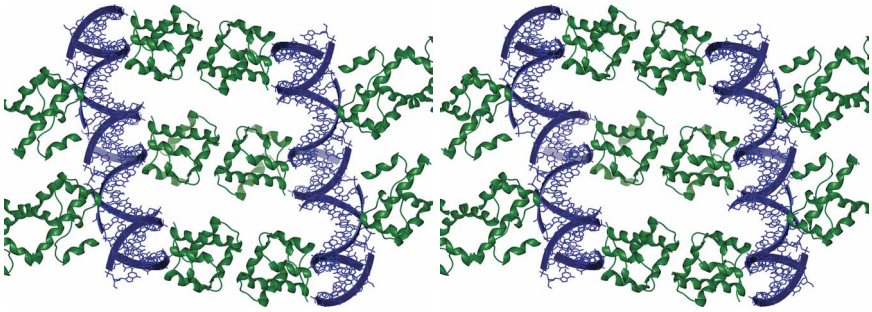

Figure 2 Model for DNA condensation by BAF. Stereo view of the packing of the BAF–7-mer DNA crystals, suggesting a model in which the protein dimers (green) simply bridge distant regions of DNA (blue).

that the previously reported discrete hexameric stoichiometry and behavior of the BAF–21-bp DNA complex[7] is a particular feature of this length of DNA, not a general property of BAF–DNA complexes.

In the crystal structure of the BAF–7-bp DNA complex, the DNA duplexes are bound at opposite ends of the BAF dimer and are approximately perpendicular to each other (**Fig. 1a**, and **Supplementary Fig. 4** and **Supplementary Table 2** online). BAF forms contacts only with the phosphate backbone of DNA on the minor groove face. The interaction is mediated by an HhH motif (**Fig. 1a**, green), a related pseudo HhH motif (**Fig. 1a**, blue) and the N terminus of α-helix 1 (α1; **Fig. 1a**, yellow) in each monomer. This mode of DNA binding is generally consistent with that predicted in ref. 10 but not with the model in ref. 11. BAF does not change conformation upon binding DNA: the r.m.s. deviation between the DNA-bound and DNA-unbound structures is 0.35 Å for all atoms. The BAF–7-bp DNA structure is consistent with previously reported mutagenesis data[10].

The HhH motif forms extensive contacts with one strand of the DNA duplex (**Fig. 1b**, green). The amide nitrogens of Gly25, Gly27, Val29 and Leu30 hydrogen bond to the phosphates of two adjacent nucleotides (5 and 6). The side chain of Val29 forms a hydrophobic interaction with the ribose of nucleotide 5. The ε-amino group of Lys6 sits in a pocket formed by the carbonyl groups of Gly21, Ile26 and Leu23, a site usually occupied by a monovalent metal ion in other HhH protein–DNA complexes[14–16], and forms a hydrogen bond with the phosphate of nucleotide 6. In contrast, the pseudo HhH motif (**Fig. 1b**, blue) forms few interactions with the complementary strand of DNA: there is a hydrogen bond between the main chain amide group of Ala71 and the phosphate of nucleotide 1′, and there is a long (over 3.5 Å) hydrogen bond between the guanidine group of Arg75 and the phosphate of nucleotide 2′ (nucleotides are numbered 1 to 7 in the 5′→3′ direction of strand 1 and 1′ to 7′ in the 5′→3′ direction of strand 2).

The binding of DNA is also aided by the N terminus of α1 (**Fig. 1b**, yellow), which forms hydrogen bonds between the main chain amide groups of Gln5 and Lys6 and the last phosphate (nucleotide 7) of the strand. In addition, the helix dipole of α1 points directly to this phosphate and the side chain of Ser4 forms a water-mediated hydrogen bond with the phosphate of nucleotide 6. These contacts represent previously unobserved structural enhancements to the HhH motif. Lys6 may enhance DNA binding in particular, as its side chain occupies the metal ion–binding site in both the DNA-bound and DNA-unbound states. In contrast, for other HhH proteins such as endonuclease III[15], AlkA[14] and MutY[16], a metal ion is bound only in the presence of DNA. By replacing the metal with the side chain of Lys6, BAF avoids the entropic cost of recruiting a metal ion from bulk solvent to assist DNA binding.

DNA binding by BAF is achieved by a very small number of contacts: only three phosphates on one strand and two on the other bind to a pair of HhH motifs in the BAF monomer. One of the pair, the pseudo HhH motif, forms at most two hydrogen bonds with DNA, leaving the majority of the interaction to the HhH motif and the supplementary contacts in α1. This minimalist, asymmetric binding mode involves no contacts to the DNA bases and uses only four protein side chains, thus ensuring sequence independence.

The packing of the BAF–7-bp DNA crystal suggests a model for the interaction of BAF with longer DNA (**Fig. 2**). In this model, both ends of the BAF dimer, along its long axis, latch onto the DNA backbone, forming a bridge between potentially distant regions of the nucleic acid chain. Subsequent binding and bridging by additional BAF dimers to other regions of the DNA collapse the DNA chain or colocalize independent DNA molecules. Thus, a structurally simple bridging capability enables BAF to compact or colocalize DNA. BAF's roles in retroviral integration, chromosomal organization, gene expression and nuclear assembly are probably related to this capability.

Accession Codes. Protein Data Bank: Coordinates have been deposited with accession code 2BZF. BIND identifiers (http://bind.ca): 316941 and 316943.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank A. Burgess Hickman, Y. Suzuki, M. Li and T. Chiu for helpful discussions. C.B. is a Paul Sigler/Argoron Fellow of the Helen Hay Whitney Foundation. This study used the helix computational systems at the Center for Information Technology, NIH, and was supported by the NIH Intramural AIDS Targeted Antiviral Program.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Brown, P.O. Integration. in *Retroviruses* (eds. Coffin, J.M., Hughes, S.H. & Varmus, H.E.) 161–203 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA, 1997).
2. Lin, C.W. & Engelman, A. *J. Virol.* **77**, 5030–5036 (2003).
3. Suzuki, Y. & Craigie, R. *J. Virol.* **76**, 12376–12380 (2002).
4. Chen, H. & Engelman, A. *Proc. Natl. Acad. Sci. USA* **95**, 15270–15274 (1998).
5. Lee, M.S. & Craigie, R. *Proc. Natl. Acad. Sci. USA* **95**, 1528–1533 (1998).
6. Segura-Totten, M. & Wilson, K.L. *Trends Cell Biol.* **14**, 261–266 (2004).
7. Zheng, R. *et al. Proc. Natl. Acad. Sci. USA* **97**, 8997–9002 (2000).
8. Margalit, A., Segura-Totten, M., Gruenbaum, Y. & Wilson, K.L. *Proc. Natl. Acad. Sci. USA* **102**, 3290–3295 (2005).
9. Furukawa, K. *et al. J. Cell Sci.* **116**, 3811–3823 (2003).
10. Umland, T.C., Wei, S.Q., Craigie, R. & Davies, D.R. *Biochemistry* **39**, 9130–9138 (2000).
11. Cai, M. *et al. Nat. Struct. Biol.* **5**, 903–909 (1998).
12. Doherty, A.J., Serpell, L.C. & Ponting, C.P. *Nucleic Acids Res.* **24**, 2488–2497 (1996).
13. Shao, X. & Grishin, N.V. *Nucleic Acids Res.* **28**, 2643–2650 (2000).
14. Hollis, T., Ichikawa, Y. & Ellenberger, T. *EMBO J.* **19**, 758–766 (2000).
15. Fromme, J.C. & Verdine, G.L. *EMBO J.* **22**, 3461–3471 (2003).
16. Fromme, J.C., Banerjee, A., Huang, S.J. & Verdine, G.L. *Nature* **427**, 652–656 (2004).